Fifteen Years Is a Long Time in SEO

I’ve been in an introspective mood lately.

Earlier this year (15 years after starting Distilled in 2005), we spun out a new company called SearchPilot to focus on our SEO A/B testing and meta-CMS technology (previously known as Distilled ODN), and merged the consulting and conferences part of the business with Brainlabs.

I’m now CEO of SearchPilot (which is primarily owned by the shareholders of Distilled), and am also SEO Partner at Brainlabs, so… I’m sorry everyone, but I’m very much staying in the SEO industry.

As such, it feels a bit like the end of a chapter for me rather than the end of the book, but it has still had me looking back over what’s changed and what hasn’t over the last 15 years I’ve been in the industry.

I can’t lay claim to being one of the first generation of SEO experts, but having been building websites since around 1996 and having seen the growth of Google from the beginning, I feel like maybe I’m second generation, and maybe I have some interesting stories to share with those who are newer to the game.

I’ve racked my brain to try and remember what felt significant at the time, and also looked back over the big trends through my time in the industry, to put together what I think makes an interesting reading list that most people working on the web today would do well to know about.

The big eras of search

I joked at the beginning of a presentation I gave in 2018 that the big eras of search oscillated between directives from the search engines and search engines rapidly backing away from those directives when they saw what webmasters actually did:

While that slide was a bit tongue-in-cheek, I do think that there's something to thinking about the eras like:

1. Build websites: Do you have a website? Would you like a website? It's hard to believe now, but in the early days of the web, a lot of folks needed to be persuaded to get their business online at all.
2. Keywords: Basic information retrieval became adversarial information retrieval as webmasters realized that they

could game the system with keyword stuffing, hidden text, and more.
3. Links: As the scale of the web grew beyond user-curated directories, link-based algorithms for search began to dominate.
4. Not those links: Link-based algorithms began to give way to adversarial link-based algorithms as webmasters swapped, bought, and manipulated links across the web graph.
5. Content for the long tail: Alongside this era, the length of the long tail began to be better-understood by both webmasters and by Google themselves – and it was in the interest of both parties to create massive amounts of (often obscure) content and get it indexed for when it was needed.
6. Not that content: Perhaps predictably (see the trend here?), the average quality of content returned in search results dropped dramatically, and so we see the first machine learning ranking factors in the form of attempts to assess "quality" (alongside relevance and website authority).
7. Machine learning: Arguably everything from that point onwards has been an adventure into machine learning and artificial intelligence, and has also taken place during the careers of most marketers working in SEO today. So, while I love writing about that stuff, I'll return to it another day.

History of SEO: crucial moments

Although I'm sure that there are interesting stories to be told about the pre-Google era of SEO, I'm not the right person to tell them (if you have a great resource, please do drop it in the comments), so let's start early in the Google journey:

Google’s foundational technology

Even if you’re coming into SEO in 2020, in a world of machine-learned ranking factors, I’d still recommend going back and reading the surprisingly accessible early academic work:

If you weren’t using the web back then, it’s probably hard to imagine what a step-change improvement Google’s PageRank-based algorithm was over the “state-of-the-art” at the time (and it’s hard to remember, even for those of us that were):

Google’s IPO

In more “things that are hard to remember clearly,” at the time of Google’s IPO in 2004, very few people expected Google to become one of the most profitable companies ever. In the early days, the founders had talked of their disdain for advertising, and had experimented with keyword-based adverts somewhat reluctantly. Because of this attitude, even within the company, most employees didn’t know what a rocket ship they were building.

From this era, I’d recommend reading the founders’ IPO letter (see this great article from Danny Sullivan – who’s ironically

now @SearchLiaison at Google):

“Our search results are the best we know how to produce. They are unbiased and objective, and we do not accept payment for them or for inclusion or more frequent updating.”

“Because we do not charge merchants for inclusion in Froogle [now Google shopping], our users can browse product categories or conduct product searches with confidence that the results we provide are relevant and unbiased.” – S1 Filing

In addition, In the Plex is an enjoyable book published in 2011 by Steven Levy. It tells the story of what then-CEO Eric Schmidt called (around the time of the IPO) “the hiding strategy”:

“Those who knew the secret … were instructed quite firmly to keep their mouths shut about it.”

“What Google was hiding was how it had cracked the code to making money on the Internet.”

Luckily for Google, for users, and even for organic search marketers, it turned out that this wasn’t actually incompatible with their pure ideals from the pre-IPO days because, as Levy recounts, “in repeated tests, searchers were happier with pages with ads than those where they were suppressed”. Phew!

Index everything

In April 2003, Google acquired a company called Applied Semantics and set in motion a series of events that I think might be the most underrated part of Google’s history.

Applied Semantics technology was integrated with their own contextual ad technology to form what became AdSense. Although the revenue from AdSense has always been dwarfed by AdWords (now just “Google Ads”), its importance in the history of SEO

is hard to understate.

By democratizing the monetization of content on the web and enabling everyone to get paid for producing obscure content, it funded the creation of absurd amounts of that content.

Most of this content would have never been seen if it weren't for the existence of a search engine that excelled in its ability to deliver great results for long tail searches, even if those searches were incredibly infrequent or had never been seen before.

In this way, Google's search engine (and search advertising business) formed a powerful flywheel with its AdSense business, enabling the funding of the content creation it needed to differentiate itself with the largest and most complete index of the web.

As with so many chapters in the story, though, it also created a monster in the form of low quality or even auto-generated content that would ultimately lead to PR crises and massive efforts to fix.

If you're interested in the index everything era, you can read more of my thoughts about it in slide 47+ of From the Horse's Mouth.

Web spam

The first forms of spam on the internet were various forms of messages, which hit the mainstream as email spam. During the early 2000s, Google started talking about the problem they'd ultimately term "web spam" (the earliest mention I've seen of link spam is in an Amit Singhal presentation from 2005 entitled Challenges in running a Commercial Web Search Engine [PDF]).

I suspect that even people who start in SEO today might've heard of Matt Cutts – the first head of webspam – as he's

still referenced often despite not having worked at Google since 2014. I enjoyed this 2015 presentation that talks about his career trajectory at Google.

Search quality era

Over time, as a result of the opposing nature of webmasters trying to make money versus Google (and others) trying to make the best search engine they could, pure web spam wasn't the only quality problem Google was facing. The cat-and-mouse game of spotting manipulation – particularly of on-page content, external links, and anchor text) – would be a defining feature of the next decade-plus of search.

It was after Singhal's presentation above that Eric Schmidt (then Google's CEO) said, "Brands are the solution, not the problem… Brands are how you sort out the cesspool".

Those who are newer to the industry will likely have experienced some Google updates (such as recent "core updates") first-hand, and have quite likely heard of a few specific older updates. But "Vince", which came after "Florida" (the first major confirmed Google update), and rolled out shortly after Schmidt's pronouncements on brand, was a particularly notable one for favoring big brands. If you haven't followed all the history, you can read up on key past updates here:

A real reputational threat

As I mentioned above in the AdSense section, there were strong incentives for webmasters to create tons of content, thus targeting the blossoming long tail of search. If you had a strong enough domain, Google would crawl and index immense numbers of pages, and for obscure enough queries, any matching content would potentially rank. This triggered the rapid growth of so-called “content farms” that mined keyword data from anywhere they could, and spun out low-quality keyword-matching content. At the same time, websites were succeeding by allowing large databases of content to get indexed even as very thin pages, or by allowing huge numbers of pages of user-generated content to get indexed.

This was a real reputational threat to Google, and broke out of the search and SEO echo chamber. It had become such a bugbear of communities like Hacker News and StackOverflow, that Matt Cutts submitted a personal update to the Hacker News community when Google launched an update targeted at fixing one specific symptom – namely that scraper websites were routinely outranking the original content they were copying.

Shortly afterwards, Google rolled out the update initially

named the “farmer update”. After it launched, we learned it had been made possible because of a breakthrough by an engineer called Panda, hence it was called the “big Panda” update internally at Google, and since then the SEO community has mainly called it the Panda update.

Although we speculated that the internal working of the update was one of the first real uses of machine learning in the core of the organic search algorithm at Google, the features it was modelling were more easily understood as human-centric quality factors, and so we began recommending SEO-targeted changes to our clients based on the results of human quality surveys.

Everything goes mobile-first

I gave a presentation at SearchLove London in 2014 where I talked about the unbelievable growth and scale of mobile and about how late we were to realizing quite how seriously Google was taking this. I highlighted the surprise many felt hearing that Google was designing mobile first:

“Towards the end of last year we launched some pretty big design improvements for search on mobile and tablet devices. Today we’ve carried over several of those changes to the desktop experience.” – Jon Wiley (lead engineer for Google Search speaking on Google+, which means there’s nowhere to link to as a perfect reference for the quote but it’s referenced here as well as in my presentation).

This surprise came despite the fact that, by the time I gave this presentation in 2014, we knew that mobile search had begun to cannibalize desktop search (and we’d seen the first drop in desktop search volumes):

And it came even though people were starting to say that the first year of Google making the majority of its revenue on mobile was less than two years away:

Writing this in 2020, it feels as though we have fully internalized how big a deal mobile is, but it’s interesting to remember that it took a while for it to sink in.

Machine learning becomes the norm

Since the Panda update, machine learning was mentioned more

and more in the official communications from Google about algorithm updates, and it was implicated in even more. We know that, historically, there had been resistance from some quarters (including from Singhal) towards using machine learning in the core algorithm due to the way it prevented human engineers from explaining the results. In 2015, Sundar Pichai took over as CEO, moved Singhal aside (though this may have been for other reasons), and installed AI / ML fans in key roles.

It goes full-circle

Back before the Florida update (in fact, until Google rolled out an update they called Fritz in the summer of 2003), search results used to shuffle regularly in a process nicknamed the Google Dance:

Most things have been moving more real-time ever since, but recent “Core Updates” appear to have brought back this kind of dynamic where changes happen on Google’s schedule rather than based on the timelines of website changes. I’ve speculated that this is because “core updates” are really Google retraining a massive deep learning model that is very customized to the shape of the web at the time. Whatever the

cause, our experience working with a wide range of clients is consistent with the official line from Google that:

Broad core updates tend to happen every few months. Content that was impacted by one might not recover – assuming improvements have been made – until the next broad core update is released.

Tying recent trends and discoveries like this back to ancient history like the Google Dance is just one of the ways in which knowing the history of SEO is “useful”.

If you’re interested in all this

I hope this journey through my memories has been interesting. For those of you who also worked in the industry through these years, what did I miss? What are the really big milestones you remember? Drop them in the comments below or hit me up on Twitter.

If you liked this walk down memory lane, you might also like my presentation From the Horse’s Mouth, where I attempt to use official and unofficial Google statements to unpack what is really going on behind the scenes, and try to give some tips for doing the same yourself:

To help us serve you better, please consider taking the 2020 Moz Blog Reader Survey, which asks about who you are, what challenges you face, and what you’d like to see more of on the Moz Blog.

Take the Survey

https://feedpress.me/link/9375/13814660/15-years-in-seo